arXiv:0807.4753v1 [quant-ph] 30 Jul 2008

Counterexamples to the maximal p-norm multiplicativity conjecture for all $\mathrm{p} > 1$

Patrick Hayden[1,*] and Andreas Winter[2,3,†]

[1] *School of Computer Science, McGill University, Montreal, Quebec, H3A 2A7, Canada*
[2] *Department of Mathematics, University of Bristol, University Walk, Bristol BS8 1TW, U. K.*
[3] *Centre for Quantum Technologies, National University of Singapore, 2 Science Drive 3, Singapore 117542*

(Dated: 30 July 2008)

For all $p > 1$, we demonstrate the existence of quantum channels with non-multiplicative maximal output p-norms. Equivalently, for all $p > 1$, the minimum output Rényi entropy of order p of a quantum channel is not additive. The violations found are large; in all cases, the minimum output Rényi entropy of order p for a product channel need not be significantly greater than the minimum output entropy of its individual factors. Since $p = 1$ corresponds to the von Neumann entropy, these counterexamples demonstrate that if the additivity conjecture of quantum information theory is true, it cannot be proved as a consequence of any channel-independent guarantee of maximal p-norm multiplicativity. We also show that a class of channels previously studied in the context of approximate encryption lead to counterexamples for all $p > 2$.

I. INTRODUCTION

The oldest problem of quantum information theory is arguably to determine the capacity of a quantum-mechanical communications channel for carrying information, specifically *classical* bits of information. (Until the 1990's it would have been unnecessary to add that additional qualification, but today the field is equally concerned with other forms of information like *qubits* and *ebits* that are fundamentally quantum-mechanical.) The classical capacity problem long predates the invention of quantum source coding [1, 2] and was of concern to the founders of information theory themselves [3]. The first major result on the problem came with the resolution of a conjecture of Gordon's [4] by Alexander Holevo in 1973, when he published the first proof [5] that the maximum amount of information that can be extracted from an ensemble of states σ_i occurring with probabilities p_i is bounded above by

$$\chi(\{p_i, \sigma_i\}) = H\Big(\sum_i p_i \sigma_i\Big) - \sum_i p_i H(\sigma_i), \tag{1}$$

where $H(\sigma) = -\operatorname{Tr} \sigma \ln \sigma$ is the von Neumann entropy of the density operator σ. For a quantum channel $\mathcal{N}$, one can then define the Holevo capacity

$$\chi(\mathcal{N}) = \max\nolimits_{\{p_i, \rho_i\}} \chi(\{p_i, \mathcal{N}(\rho_i)\}), \tag{2}$$

where the maximization is over all ensembles of input states. Writing $C(\mathcal{N})$ for the classical capacity of the channel $\mathcal{N}$, this leads easily to an upper bound of

$$C(\mathcal{N}) \leq \lim_{n \to \infty} \frac{1}{n} \chi(\mathcal{N}^{\otimes n}). \tag{3}$$

[*] Electronic address: patrick@cs.mcgill.ca
[†] Electronic address: a.j.winter@bris.ac.uk

It then took more than two decades for further substantial progress to be made on the problem, but in 1996, building on recent advances [6], Holevo [7] and Schumacher-Westmoreland [8] managed to show that the upper bound in Eq. (3) is actually achieved. This was a resolution of sorts to the capacity problem, but the limit in the equation makes it in practice extremely difficult to evaluate. If the codewords used for data transmission are restricted such that they are not entangled across multiple uses of the channel, however, the resulting *product state capacity* $C_{1\infty}(\mathcal{N})$ has the simpler expression

$$C_{1\infty}(\mathcal{N}) = \chi(\mathcal{N}). \tag{4}$$

The additivity conjecture for the Holevo capacity asserts that for all channels $\mathcal{N}_1$ and $\mathcal{N}_2$,

$$\chi(\mathcal{N}_1 \otimes \mathcal{N}_2) = \chi(\mathcal{N}_1) + \chi(\mathcal{N}_2). \tag{5}$$

This would imply, in particular, that $C_{1\infty}(\mathcal{N}) = C(\mathcal{N})$, or that entangled codewords do not increase the classical capacity of a quantum channel.

In 2003, Peter Shor [9], building on several previously established connections [10, 11, 12], demonstrated that the additivity of the Holevo capacity, the additivity of the entanglement of formation [13, 14, 15, 16] and the superadditivity of the entanglement of formation [17] are all equivalent to another conjecture of Shor's which is particularly simple to express mathematically, known as the *minimum output entropy conjecture* [18]. For a channel $\mathcal{N}$, define

$$H^{\min}(\mathcal{N}) = \min_{|\varphi\rangle} H(\mathcal{N}(\varphi)), \tag{6}$$

where the minimization is over all pure input states $|\varphi\rangle$. The minimum output entropy conjecture asserts that for all channels $\mathcal{N}_1$ and $\mathcal{N}_2$,

$$H^{\min}(\mathcal{N}_1 \otimes \mathcal{N}_2) = H^{\min}(\mathcal{N}_1) + H^{\min}(\mathcal{N}_2). \tag{7}$$

There has been a great deal of previous work on these conjectures, particularly inconclusive numerical searches for counterexamples, necessarily in low dimension, at Caltech, IBM, in Braunschweig (IMaPh) and Tokyo (ERATO) [19], as well as proofs of many special cases. For example, the minimum output entropy conjecture has been shown to hold if one of the channels is the identity channel [20, 21], a unital qubit channel [22], a generalized depolarizing channel [23, 24] or an entanglement-breaking channel [25, 26, 27]. In addition, the weak additivity conjecture was confirmed for generalized dephasing channels [28], the conjugates of all these channels [29] and some other special classes of channels [16, 30, 31, 32]. Further evidence for qubit channels was supplied in [18]. This list is by no means exhaustive. The reader is directed to Holevo's reviews for a detailed account of the history of the additivity problem [33, 34].

For the past several years, the most commonly used strategy for proving these partial results has been to demonstrate the multiplicativity of maximal p-norms of quantum channels for p approaching 1 [20]. For a quantum channel $\mathcal{N}$ and $p > 1$, define the maximal p-norm of $\mathcal{N}$ to be

$$\nu_p(\mathcal{N}) = \sup\Big\{ \big\|\mathcal{N}(\rho)\big\|_p ; \rho \geq 0,\ \operatorname{Tr}\rho = 1 \Big\}. \tag{8}$$

In the equation, $\|\sigma\|_p = \big(\operatorname{Tr}|\sigma|^p\big)^{1/p}$. The *maximal p-norm multiplicativity conjecture* [20] asserts that for all quantum channels $\mathcal{N}_1$ and $\mathcal{N}_2$,

$$\nu_p(\mathcal{N}_1 \otimes \mathcal{N}_2) = \nu_p(\mathcal{N}_1)\nu_p(\mathcal{N}_2). \tag{9}$$

This can be re-expressed in an equivalent form more convenient to us using Rényi entropies. Define the Rényi entropy of order p to be

$$H_p(\rho) = \frac{1}{1-p} \ln \operatorname{Tr} \rho^p \tag{10}$$

for $p > 0$, $p \neq 1$. Since $\lim_{p \downarrow 1} H_p(\rho) = H(\rho)$, we will also define $H_1(\rho)$ to be $H(\rho)$. All these entropies have the property that they are 0 for pure states and achieve their maximum value of the logarithm of the dimension on maximally mixed states. Define the minimum output Rényi entropy $H_p^{\min}$ by substituting H_p for H in Eq. (6). Since $H_p^{\min}(\mathcal{N}) = \frac{p}{1-p} \ln \nu_p(\mathcal{N})$, Eq. (9) can then be written equivalently as

$$H_p^{\min}(\mathcal{N}_1 \otimes \mathcal{N}_2) = H_p^{\min}(\mathcal{N}_1) + H_p^{\min}(\mathcal{N}_2), \tag{11}$$

in which form it is clear that the maximal p-norm multiplicativity conjecture is a natural strengthening of the original minimum output entropy conjecture (7).

This conjecture spawned a significant literature of its own which we will not attempt to summarize. Holevo's reviews are again an excellent source [33, 34]. Some more recent important references include [35, 36, 37, 38, 39, 40]. Unlike the von Neumann entropy case, however, some counterexamples had already been found prior to this paper. Namely, Werner and Holevo found a counterexample to Eq. (11) for $p > 4.79$ [41] that nonetheless doesn't violate the p-norm multiplicativity conjecture for $1 < p < 2$ [42, 43, 44].

Moreover, in 2007, Winter showed that a class of channels that had previously been studied in the context of approximate encryption provide counterexamples to the conjecture for all $p > 2$ [45]. In light of these developments, the standing conjecture was that the maximal p-norm multiplicativity held for $1 \leq p \leq 2$, corresponding to the region in which the map $X \mapsto X^p$ is operator convex [35]. More conservatively, it was conjectured to hold at least in an open interval $(1, 1+\epsilon)$, which would be sufficient to imply the minimum output entropy conjecture. On the contrary, shortly after Winter's discovery, Hayden showed that the conjecture is false for all $1 < p < 2$ [46].

The current paper merges and slightly strengthens [45] and [46]. We begin in Section II, by presenting Winter's counterexamples from [45], which share some important features with [46] but are simpler to analyze. Section III then presents Hayden's counterexamples from [46] with an improved analysis showing that they work for all $p > 1$, not just $1 < p < 2$.

In particular, given $p > 1$, we show that there exist channels $\mathcal{N}_1$ and $\mathcal{N}_2$ with output dimension d such that both $H_p^{\min}(\mathcal{N}_1)$ and $H_p^{\min}(\mathcal{N}_2)$ are equal to $\ln d - \mathcal{O}(1)$ but $H_p^{\min}(\mathcal{N}_1 \otimes \mathcal{N}_2) = \ln d + \mathcal{O}(1)$, so

$$H_p^{\min}(\mathcal{N}_1) + H_p^{\min}(\mathcal{N}_2) - H_p^{\min}(\mathcal{N}_1 \otimes \mathcal{N}_2) = \ln d - \mathcal{O}(1). \tag{12}$$

Thus, one finds that the minimum output entropy of the product channel need not be significantly larger than the minimum output entropy of the individual factors. Since [20, 24]

$$H_p^{\min}(\mathcal{N}_1 \otimes \mathcal{N}_2) \geq H_p^{\min}(\mathcal{N}_1) = \ln d - \mathcal{O}(1), \tag{13}$$

these counterexamples are essentially the strongest possible for all $p > 1$, up to a constant additive term. (Note that the dependence of $H_p^{\min}$ on p is absorbed here in the asymptotic notation.)

At $p = 1$ itself, however, we see no evidence of a violation of the additivity conjecture for the channels we study. Thus, the conjecture stands and it is still an open question whether entangled codewords can increase the classical capacity of a quantum channel.

Notation: If A and B are finite dimensional Hilbert spaces, we write $AB \equiv A \otimes B$ for their tensor product and $|A|$ for $\dim A$. The Hilbert spaces on which linear operators act will be denoted by a superscript. For instance, we write φ^{AB} for a density operator on AB. Partial traces will be abbreviated by omitting superscripts, such as $\varphi^A \equiv \operatorname{Tr}_B \varphi^{AB}$. We use a similar notation for pure states, e.g. $|\psi\rangle^{AB} \in AB$, while abbreviating $\psi^{AB} \equiv |\psi\rangle\langle\psi|^{AB}$. We associate to any two isomorphic Hilbert spaces $A \simeq A'$ a unique maximally entangled state which we denote $|\Phi\rangle^{AA'}$. Given any orthonormal basis $\{|i\rangle^A\}$ for A, if we define $|i\rangle^{A'} = V|i\rangle^A$ where V is the associated isomorphism, we can write this state as $|\Phi\rangle^{AA'} = |A|^{-1/2}\sum_{i=1}^{|A|}|i\rangle^A|i\rangle^{A'}$. We will also make use of the asymptotic notation $f(n) = \mathcal{O}(g(n))$ if there exists $C > 0$ such that for sufficiently large n, $|f(n)| \leq Cg(n)$. $f(n) = \Omega(g(n))$ is defined similarly but with the reverse inequality $|f(n)| \geq Cg(n)$. Finally, $f(n) = \Theta(g(n))$ if $f(n) = \mathcal{O}(g(n))$ and $f(n) = \Omega(g(n))$.

II. RANDOM UNITARY CHANNELS: $\mathbf{p > 2}$

This class of counterexamples, while only working for $p > 2$, has the advantage of being a straightforward application of well-known results. Later in the paper we will present stronger counterexamples that reuse the same basic strategy, albeit with some additional technical complications. A random unitary channel is a map of the form

$$\mathcal{N} : \rho \longmapsto \frac{1}{n}\sum_{i=1}^{n} V_i \rho V_i^\dagger, \tag{14}$$

with the V_i unitary transformations of an underlying (finite dimensional) Hilbert space. Let d be the dimension of this space. Following [47], we call $\mathcal{N}$ *ϵ-randomizing* if for all ρ,

$$\left\| \mathcal{N}(\rho) - \frac{1}{d}I \right\|_\infty \leq \frac{\epsilon}{d}. \tag{15}$$

In that paper, it was shown that for $0 < \epsilon < 1$, ϵ-randomizing channels exist in all dimensions $d > \frac{10}{\epsilon}$, with $n = \frac{134}{\epsilon^2} d \ln d$. In fact, randomly picking the V_i from the Haar measure on the unitary group will, with high probability, yield such a channel.

Recently, it was shown by Aubrun [48] that n can in fact be taken to be $\mathcal{O}(d/\epsilon^2)$ for Haar distributed V_i, and $\mathcal{O}(d(\ln d)^4/\epsilon^2)$ for V_i drawn from any ensemble of exactly randomizing unitaries.

Lemma II.1 *For a random unitary channel $\mathcal{N}$ and its complex conjugate, $\overline{\mathcal{N}} : \rho \mapsto \frac{1}{n}\sum \overline{V_i}\rho\overline{V_i}^\dagger$, one has $\nu_p(\mathcal{N} \otimes \overline{\mathcal{N}}) \geq \frac{1}{n}$.*

Proof We use the maximally entangled state $|\Phi\rangle = d^{-1/2}\sum_i |i\rangle|i\rangle$ as test state, abbreviating $\Phi = |\Phi\rangle\langle\Phi|$:

$$\begin{aligned}
\nu_p(\mathcal{N} \otimes \overline{\mathcal{N}}) &\geq \left\| (\mathcal{N} \otimes \overline{\mathcal{N}})\Phi \right\|_p \\
&= \left\| \frac{1}{n^2}\sum_{i,j=1}^{n} (V_i \otimes \overline{V_j})\Phi(V_i \otimes \overline{V_j})^\dagger \right\|_p \\
&= \left\| \frac{1}{n}\Phi + \frac{1}{n^2}\sum_{i\neq j}(V_i \otimes \overline{V_j})\Phi(V_i \otimes \overline{V_j})^\dagger \right\|_p \geq \frac{1}{n},
\end{aligned}$$

where in the third line we have invoked the $U \otimes \overline{U}$-invariance of Φ for the n terms when $i = j$. For the final inequality, observe that the largest eigenvalue λ_1 of $(\mathcal{N} \otimes \overline{\mathcal{N}})\Phi$ is at least $\frac{1}{n}$. Denoting the other eigenvalues λ_α, $\|(\mathcal{N} \otimes \overline{\mathcal{N}})\Phi\|_p = (\sum_\alpha \lambda_\alpha^p)^{1/p} \geq \lambda_1$, and we are done. □

Lemma II.2 *If the channel $\mathcal{N}$ is ϵ-randomizing, then when $p > 1$,*

$$\nu_p(\mathcal{N}) = \nu_p(\overline{\mathcal{N}}) \leq \left(\frac{1+\epsilon}{d}\right)^{1-1/p}.$$

Proof Clearly, $\mathcal{N}$ and $\overline{\mathcal{N}}$ have the same maximum output p-norm. For the former, observe that the ϵ-randomizing condition implies that for an arbitrary input state ρ, $\|\mathcal{N}(\rho)\|_\infty \leq \frac{1+\epsilon}{d}$. In other words, all the eigenvalues λ_α of the output state $\mathcal{N}(\rho)$ are bounded between 0 and $\frac{1+\epsilon}{d}$. In addition, because $\mathcal{N}(\rho)$ is a density operator, the eigenvalues sum to 1.

Subject to these constraints, however, the convexity of the function $x \mapsto x^p$ ensures that the p-norm $\|\mathcal{N}(\rho)\|_p = (\sum_\alpha \lambda_\alpha^p)^{1/p}$ is maximized when the largest eigenvalue is $\frac{1+\epsilon}{d}$ and it occurs with multiplicity $\lfloor \frac{d}{1+\epsilon} \rfloor$, and all but possibly one remaining eigenvalue is 0. Thus,

$$\|\mathcal{N}(\rho)\|_p = \left(\sum_\alpha \lambda_\alpha^p\right)^{1/p} \leq \left(\frac{d}{1+\epsilon}\left(\frac{1+\epsilon}{d}\right)^p\right)^{1/p} = \left(\frac{1+\epsilon}{d}\right)^{1-1/p}. \tag{16}$$

□

Theorem II.3 *Fix any $0 < \epsilon < 1$ and a family of ϵ-randomizing maps $\mathcal{N}$ as in Eq. (14) with $n > 134\, d \ln d/\epsilon^2$. Then, for any $p > 2$ and sufficiently large d,*

$$\nu_p(\mathcal{N})\nu_p(\overline{\mathcal{N}}) \leq \left(\frac{1+\epsilon}{d}\right)^{2-2/p} \ll \frac{1}{n} \leq \nu_p(\mathcal{N} \otimes \overline{\mathcal{N}}), \tag{17}$$

In other words, for this family of channels, the maximum output p-norm is strictly supermultiplicative for sufficiently large d when $p > 2$.

Proof Follows from Lemmas II.1 and II.2 since $2 - 2/p > 1$. □

These counterexamples to the multiplicativity of the output p-norm for $p > 2$ are interesting in that they are random unitary channels, which are among the simplest truly quantum maps. In fact, the first proofs of multiplicativity for unital qubit channels [22] and depolarizing channels [24] exploited this type of structure. Indeed, unital qubit channels are always random unitary channels (with $d = 2$) [18]. Despite the fact that King showed multiplicativity for such channels at all $p > 1$ [22], there is no conflict with the result here, as the bound on n becomes better than d^2 only for rather large dimension d.

We observe, furthermore, that $p = 2$ is indeed the limit of validity of this class of counterexamples, since $n \geq d$ for any ϵ-randomizing map.

III. GENERIC QUANTUM CHANNELS: ALL $\mathbf{p > 1}$

Let E, F and G be finite dimensional quantum systems, then define $R = E$, $S = FG$, $A = EF$ and $B = G$, so that $RS = AB = EFG$. Our second and stronger class of counterexamples will be channels from S to A of the form

$$\mathcal{N}(\rho) = \mathrm{Tr}_B \left[U(|0\rangle\langle 0|^R \otimes \rho)U^\dagger\right] \tag{18}$$

for U unitary and $|0\rangle$ some fixed state on R. Another, slightly more flexible way of writing this is in the language of isometric Stinespring dilations: namely, the Hilbert space isometry $V : S \hookrightarrow AB$ defined by $V|\varphi\rangle = U\big(|0\rangle^R|\varphi\rangle^S\big)$. In this notation, to which we will adhere from now, $\mathcal{N}(\rho) = \operatorname{Tr}_B V\rho V^\dagger$.

Our method will be to fix the dimensions of the systems involved, select U (i.e., the isometry V) at random, and show that the resulting channel is likely to violate additivity. The rough intuition motivating our examples is the same as in the previous section: we will exploit the fact that there are channels that appear to be highly depolarizing for product state inputs despite the fact that they are not close to the depolarizing channel in, for example, the norm of complete boundedness [49].

Consider a single copy of $\mathcal{N}$ and the associated map $V : |\varphi\rangle^S \mapsto U\big(|0\rangle^R|\varphi\rangle^S\big)$. This map takes S to a subspace of $A \otimes B$, and if U is selected according to the Haar measure, then the image of S is itself a random subspace, distributed according to the unitarily invariant measure. In [50], it was shown that if $|S|$ is chosen appropriately, then the image is likely to contain only almost maximally entangled states, as measured by the entropy of entanglement. After tracing over B, this entropy of entanglement becomes the entropy of the output state. Thus, for S of suitable size, all input states get mapped to high entropy output states. We will repeat the analysis below, finding that the maximum allowable size of S will depend on p as described by the following two lemmas.

Lemma III.1 *The maps $f_p(|\varphi\rangle) = H_p(\varphi^A)$ on unit vectors (states) $|\varphi\rangle \in A \otimes B$, $2 \leq |A| \leq |B|$, have expectation*

$$\mathbb{E}f_p \geq \mathbb{E}f_\infty \geq \ln|A| - \gamma\sqrt{|A|/|B|}, \tag{19}$$

for a uniformly random state φ, with a universal constant γ which may be chosen arbitrarily close to 3 for sufficiently large $|A|$.
Furthermore, for $p > 1$, the functions f_p are all Lipschitz continuous, with the Lipschitz constant Λ_p bounded above by

$$\Lambda_p^2 \leq \frac{4p^2}{(1-p)^2}|A|^{1-\frac{1}{p}}. \tag{20}$$

Proof The first inequality in Eq. (19) is by the monotonicity of the Rényi entropies in p. For the second, observe $f_\infty(|\varphi\rangle) = -\ln\|\varphi^A\|_\infty$, so

$$\mathbb{E}f_\infty(|\varphi\rangle) = \mathbb{E}\big(-\ln\left\|\varphi^A\right\|_\infty\big) \geq -\ln\mathbb{E}\left\|\varphi^A\right\|_\infty.$$

The expectation of the largest eigenvalue of φ^A has been widely studied in random matrix theory. Just note that $|\varphi\rangle$ is well-approximated by a Gaussian unit vector $|\Gamma\rangle$, that is, a random vector all of whose real and imaginary components (in any basis) are i.i.d. normal with expectation 0 and variance $1/2|A||B|$. (See [51, Appendix].) Indeed, by the triangle inequality,

$$\mathbb{E}_\varphi\left\|\varphi^A\right\|_\infty \leq \mathbb{E}_\Gamma\left\|\Gamma^A\right\|_\infty,$$

and the right hand side, for large A and B, is known [52, 53] to be asymptotically

$$\frac{\left(\sqrt{|A|}+\sqrt{|B|}\right)^2}{|A||B|} = \frac{1}{|A|} + \frac{2}{\sqrt{|A||B|}} + \frac{1}{|B|} \leq \frac{1}{|A|}\left(1 + 3\sqrt{\frac{|A|}{|B|}}\right).$$

The explicit upper bound of $\frac{\left(\sqrt{|A|}+\sqrt{|B|}\right)^2}{|A||B|}$ has been obtained for matrices with *real* Gaussian entries [54], but the analogous statement for complex Gaussian entries seems to be unknown.

Now, for the Lipschitz bound: we proceed as in [50], inferring the general bound from a Lipschitz bound for the Rényi entropy of a dephased version on φ^A. Fix bases $\{|j\rangle\}$ and $\{|k\rangle\}$ of A and B, respectively, so that we can write $|\varphi\rangle = \sum_{jk} \varphi_{jk}|j\rangle|k\rangle$, where the coefficients are to be decomposed into real and imaginary parts: $\varphi_{jk} = t_{jk0} + it_{jk1}$.

We actually show that

$$g_p(|\varphi\rangle) = \frac{1}{1-p}\ln \operatorname{Tr}\left[\left(\sum_j |j\rangle\langle j|\varphi^A|j\rangle\langle j|\right)^p\right] = \frac{1}{1-p}\ln\sum_j \langle j|\varphi^A|j\rangle^p = \frac{1}{1-p}\ln\sum_j\left(\sum_{kz} t_{jkz}^2\right)^p$$

is $\frac{2p}{p-1}|A|^{1/2-1/2p}$-Lipschitz. This implies the result for f_p as follows. Note first that $g_p(|\varphi\rangle) \geq f_p(|\varphi\rangle)$, with equality if $\{|j\rangle\}$ is an eigenbasis of φ^A. Now, for two vectors $|\varphi\rangle$, $|\psi\rangle$, we may without loss of generality assume that $f_p(|\psi\rangle) \geq f_p(|\varphi\rangle)$, and that $\{|j\rangle\}$ is the eigenbasis of φ^A. Thus, by assumption,

$$f_p(|\psi\rangle) - f_p(|\varphi\rangle) \leq g_p(|\psi\rangle) - g_p(|\varphi\rangle) \leq \frac{2p}{p-1}|A|^{1/2-1/2p}\||\psi\rangle - |\varphi\rangle\|_2.$$

To bound the Lipschitz constant of g_p, it is sufficient to find an upper bound on its gradient. It is straightforward to see that

$$\frac{\partial g_p}{\partial t_{jkz}} = \frac{1}{1-p}\frac{1}{\sum_{j'}\left(\sum_{k'z'} t_{j'k'z'}^2\right)^p}\cdot 2\,p\,t_{jkz}\left(\sum_{k'z'} t_{jk'z'}^2\right)^{p-1},$$

so introducing the notation $x_j = \sum_{kz} t_{jkz}^2$, we have

$$\left\|\nabla g_p\right\|_2^2 = \frac{4p^2}{(1-p)^2}\frac{\sum_j x_j^{2p-1}}{\left(\sum_j x_j^p\right)^2} = \frac{4p^2}{(1-p)^2}\frac{\sum_j (x_j^p)^{(2p-1)/p}}{\left(\sum_j x_j^p\right)^2},$$

which we need to maximize subject to the constraint $\sum_j x_j = 1$. Since $(2p-1)/p \geq 1$, the function $y^{(2p-1)/p}$ is convex. Therefore, for fixed $s = \sum_j x_j^p \geq |A|^{1-p}$, the right hand side is maximal when all the x_j^p except for one are 0. Thus,

$$\left\|\nabla g_p\right\|_2^2 \leq \max_{|A|^{1-p}\leq s\leq 1}\frac{4p^2}{(1-p)^2}s^{[(2p-1)/p]-2} = \frac{4p^2}{(1-p)^2}|A|^{1-1/p},$$

and we are done. □

Lemma III.2 *Let A and B be quantum systems with $2 \leq |A| \leq |B|$ and $1 < p \leq \infty$. Then there exists a subspace $S \subset A \otimes B$ of dimension*

$$|S| = \left\lfloor \frac{c}{4}\left(1-\frac{1}{p}\right)^2\frac{\alpha^2}{\ln(5/\delta)}|A|^{1/p}|B|\right\rfloor \tag{21}$$

(with a universal constant c), that contains only states $|\varphi\rangle \in S$ with high entanglement, in the sense that

$$H_p(\varphi^A) \geq \ln|A| - \alpha - \beta + \ln(1-\delta), \tag{22}$$

where $\beta = \gamma\sqrt{|A|/|B|}$ is as in Lemma III.1. The probability that a subspace of dimension $|S|$ chosen at random according to the unitarily invariant measure will not have this property is bounded above by

$$2\left(\frac{5}{\delta}\right)^{2|S|}\exp\left(-c\left(1-\frac{1}{p}\right)^2\alpha^2|A|^{1/p}|B|\right). \tag{23}$$

The universal constant c may be chosen to be $1/72\pi^3$.

Proof The argument is nearly identical to the proof of Theorem IV.1 in [50], but with an improvement, possible due to the fact the we are looking at a function defined via a norm. (See [55] and [48].)

First of all, by Levy's Lemma, for a function f on pure states of $A \otimes B$ with Lipschitz constant Λ, the random variable $f(|\varphi\rangle)$ for a uniformly distributed $|\varphi\rangle$ on the unit sphere in $A \otimes B$ obeys

$$\Pr\{f < \mathbb{E}f - \alpha\} \le 2\exp\left(-\frac{2}{9\pi^3}\frac{\alpha^2}{\Lambda^2}|A||B|\right).$$

(See [50, Lemma III.1] for an exposition.) We apply this to f_p, for which we have a Lipschitz bound by Lemma III.1. Furthermore, we can find a δ-net $\mathcal{M}$ of cardinality $|\mathcal{M}| \le (5/\delta)^{2|S|}$ on the unit vectors in S [50, Lemma III.6]. In other words, for each unit vector $|\varphi\rangle \in S$ there exists a $|\tilde{\varphi}\rangle \in \mathcal{M}$ such that $\||\varphi\rangle - |\tilde{\varphi}\rangle\|_2 \le \delta$. Combining the net, the Lipschitz constant and the union bound, we get

$$\Pr_S\Big\{\exists|\varphi\rangle \in \mathcal{M} \quad f_p(|\varphi\rangle) < \ln|A| - \alpha/2 - \beta\Big\} \le \left(\frac{5}{\delta}\right)^{2|S|} 2\exp\left(-\frac{2}{9\pi^3}\frac{\alpha^2(1-1/p)^2}{16|A|^{1-1/p}}|A||B|\right),$$

which is the probability inequality claimed in the theorem. Moreover, the right hand side is less than 1 if $|S|$ is chosen as stated in the theorem.

Now, assume we have a subspace S with a δ-net $\mathcal{M}$ such that

$$(\forall|\varphi\rangle \in \mathcal{M})\,(f_p(\varphi) \ge \ln|A| - \alpha - \beta)\,, \ \text{ i.e. } \ r := \max_{|\varphi\rangle\in\mathcal{M}} \|\varphi^A\|_p \le e^{-(1-1/p)(\ln|A|-\alpha-\beta)}. \tag{24}$$

Denote

$$R := \max_{|\varphi\rangle\in S \text{ unit vector}} \|\varphi^A\|_p = \max_{\rho \text{ d.o. supported on } S} \|\rho^A\|_p,$$

where the latter equality is due to the convexity of the norm. Hence, for each unit vector $|\varphi\rangle \in S$ and corresponding $|\tilde{\varphi}\rangle \in \mathcal{M}$ such that $\|\varphi - \tilde{\varphi}\|_1 \le \delta$,

$$\|\varphi^A\|_p \le \|\tilde{\varphi}^A\|_p + \|\varphi^A - \tilde{\varphi}^A\|_p \le r + \delta R,$$

where we have used triangle inequality and the trace norm bound on $\varphi - \tilde{\varphi}$. Consequently, $R \le r/(1-\delta)$, and inserting that into Eq. (24) finishes the proof. □

Consider now the product channel $\mathcal{N} \otimes \bar{\mathcal{N}}$, where $\bar{\mathcal{N}}(\rho) = \mathrm{Tr}_B\, \bar{V}\rho V^T$ is the complex conjugate of $\mathcal{N}$. We will exploit an approximate version of the symmetry used in the random unitary channel counterexamples. Fix orthonormal bases of S, A and B to be used in the definition of maximally entangled states involving these systems. (These have to be the same product bases with respect to which we define the complex conjugate.)

In the trivial case where $|S| = |A \otimes B|$, the isometry V is unitary and the identity $V \otimes \bar{V}|\Phi\rangle = (V\bar{V}^T \otimes I)|\Phi\rangle = |\Phi\rangle$ for the maximally entangled state $|\Phi\rangle^{S_1S_2}$ implies that

$$(\mathcal{N} \otimes \bar{\mathcal{N}})(|\Phi\rangle\langle\Phi|^{S_1S_2}) = \mathrm{Tr}_{B_1B_2}\left[|\Phi\rangle\langle\Phi|^{A_1A_2} \otimes |\Phi\rangle\langle\Phi|^{B_1B_2}\right] = |\Phi\rangle\langle\Phi|^{A_1A_2}. \tag{25}$$

The output of $\mathcal{N} \otimes \bar{\mathcal{N}}$ will thus be a pure state. In the general case, we will choose $|S|/|A \otimes B|$ to be large but not trivial, in which case useful bounds can still be placed on the largest eigenvalue of the output state for an input state maximally entangled between S_1 and S_2.

Lemma III.3 *Let $|\Phi\rangle^{S_1S_2}$ be a state maximally entangled between S_1 and S_2 as in the previous paragraph. Then $(\mathcal{N} \otimes \bar{\mathcal{N}})(\Phi^{S_1S_2})$ has an eigenvalue of at least $\frac{|S|}{|A||B|}$.*

Proof This is an easy calculation again exploiting the $U \otimes \bar{U}$ invariance of the maximally entangled state. Note that whereas V is an isometric embedding, $V^\dagger$ is a partial isometry. More precisely, it can be understood as a unitary $U^\dagger$ on $A \otimes B$ followed by a fixed projection P, say onto the first $|S|$ coordinates of $A \otimes B$. Now,

$$\begin{aligned}
\left\|(\mathcal{N} \otimes \bar{\mathcal{N}})|\Phi\rangle\langle\Phi|^{S_1S_2}\right\|_\infty &\geq \mathrm{Tr}\left(\left[(\mathcal{N} \otimes \bar{\mathcal{N}})|\Phi\rangle\langle\Phi|^{S_1S_2}\right]|\Phi\rangle\langle\Phi|^{A_1A_2}\right) \\
&\geq \mathrm{Tr}\left((V \otimes \bar{V})|\Phi\rangle\langle\Phi|^{S_1S_2}(V \otimes \bar{V})^\dagger(|\Phi\rangle\langle\Phi|^{A_1A_2} \otimes |\Phi\rangle\langle\Phi|^{B_1B_2})\right) \\
&= \mathrm{Tr}\left((P \otimes \bar{P})|\Phi\rangle\langle\Phi|^{S_1S_2}(P \otimes \bar{P})(U \otimes \bar{U})^\dagger(|\Phi\rangle\langle\Phi|^{A_1A_2} \otimes |\Phi\rangle\langle\Phi|^{B_1B_2})(U \otimes \bar{U}))\right) \\
&= \mathrm{Tr}\left((P \otimes \bar{P})|\Phi\rangle\langle\Phi|^{S_1S_2}(P \otimes \bar{P})(|\Phi\rangle\langle\Phi|^{A_1A_2} \otimes |\Phi\rangle\langle\Phi|^{B_1B_2})\right) = \frac{|S|}{|A||B|},
\end{aligned}$$

and we are done. □

In order to demonstrate violations of additivity, the first step is to bound the minimum output entropy from below for a single copy of the channel. Fix $1 < p \leq \infty$, let $|B| = |A|$ so that $\beta = \gamma$, set $\alpha = \delta = 1/2$, and then choose $|S|$ according to Lemma III.2. With probability approaching 1 as $|A| \to \infty$,

$$H_p^{\min}(\mathcal{N}) \geq \ln|A| - \gamma - 1/2 - \ln 2, \tag{26}$$

when the subspace S defining the channel is chosen according to the unitary invariant measure. (Since we're interested in $|A| \to \infty$, we may choose any $\gamma > 3$.) The same obviously holds for $H_p^{\min}(\bar{\mathcal{N}})$. Recall that the entropy of the uniform distribution is $\ln|A|$ so the minimum entropy is near the maximum possible. Fix a channel such that these lower bounds on $H_p^{\min}(\mathcal{N})$ and $H_p^{\min}(\bar{\mathcal{N}})$ are satisfied.

By Lemma III.3,

$$H_p\big((\mathcal{N} \otimes \bar{\mathcal{N}})(\Phi)\big) = \frac{1}{1-p}\ln\left(\sum_\alpha \lambda_\alpha^p\right) \leq \frac{1}{1-p}\ln\left(\frac{|S|}{|A||B|}\right)^p = \frac{p}{1-p}\ln\frac{|S|}{|A||B|}, \tag{27}$$

where the λ_α are the eigenvalues of $(\mathcal{N} \otimes \bar{\mathcal{N}})(\Phi)$. Substituting the value of $|S|$ from Lemma III.2 into this inequality yields

$$H_p\big((\mathcal{N} \otimes \bar{\mathcal{N}})(\Phi)\big) \leq \ln|A| + \mathcal{O}\left(1 + \frac{p}{p-1}\ln\frac{p}{p-1}\right) \leq \ln|A| + \mathcal{O}\left((1-1/p)^{-2}\right), \tag{28}$$

where the $\mathcal{O}$ notation hides only an absolute constant, independent of $|A|$ and $p > 1$. Thus, the Rényi entropy of $(\mathcal{N} \otimes \bar{\mathcal{N}})(\Phi)$ is strictly less than $H_p^{\min}(\mathcal{N}) + H_p^{\min}(\bar{\mathcal{N}}) \geq 2\ln|A| - \mathcal{O}(1)$. This is a violation of conjecture (11), with the size of the gap approaching $\ln|A| - \mathcal{O}(1)$ for large $|A|$.

Theorem III.4 *For all $1 < p \leq \infty$, there exists a quantum channel for which the inequalities (26) and (28) both hold. The inequalities are inconsistent with the maximal p-norm multiplicativity conjecture.* □

Note, however, that changing p also requires changing $|S|$ according Lemma III.2, so we have a sequence of channels violating additivity of the minimal output Rényi entropy as p decreases to 1, as opposed to a single channel doing so for every p. This prevents us from drawing conclusions about the von Neumann entropy by taking the limit $p \to 1$. Likewise, an examination of Eq. (28) reveals that we also lose control over the two-copy minimum output entropy of a fixed channel as $p \to 1$.

Another observation comes from the fact that our examples violate additivity by so much: namely that, due to Lemma III.3, the dimension of the subspace S in Lemma III.2 is essentially optimal up to constant factors (depending on p). Any stronger violations of additivity would contradict Eq. (13), the inequality $H_p^{\min}(\mathcal{N} \otimes \bar{\mathcal{N}}) \geq H_p^{\min}(\mathcal{N})$.

As an aside, it is interesting to observe that violating maximal p-norm multiplicativity has structural consequences for the channels themselves. For example, because entanglement-breaking channels do not violate multiplicativity [56], there must be states $|\psi\rangle^{S_1 S_2}$ such that $(\mathcal{N} \otimes I^{S_2})(\psi)$ is entangled, despite the fact that $\mathcal{N}$ will be a rather noisy channel. (The same conclusions apply to the maps of section II , where the conclusion takes the form that ϵ-randomizing random unitary channels need not be entanglement-breaking.)

IV. THE VON NEUMANN ENTROPY CASE

Despite the large violations found for p close to 1, the class of examples presented here do not appear to contradict the minimum output entropy conjecture for the von Neumann entropy. The reason is that the upper bound demonstrated for $H_p\big((\mathcal{N} \otimes \bar{\mathcal{N}})(\Phi)\big)$ in the previous section rested entirely on the existence of one large eigenvalue for $(\mathcal{N} \otimes \bar{\mathcal{N}})(\Phi)$. The von Neumann entropy is not as sensitive to the value of a single eigenvalue as are the Rényi entropies for $p > 1$ and, consequently, does not appear to exhibit additivity violations. With a bit of work, it is possible to make these observations more rigorous.

Lemma IV.1 *Let $|\Phi\rangle^{S_1 S_2}$ be a maximally entangled state between S_1 and S_2. Assuming that $|A| \leq |B| \leq |S|$,*

$$\int \operatorname{Tr}\left[\big((\mathcal{N} \otimes \bar{\mathcal{N}})(|\Phi\rangle\langle\Phi|)\big)^2\right] dU = \frac{|S|^2}{|A|^2|B|^2} + \mathcal{O}\left(\frac{1}{|A|^2}\right), \tag{29}$$

where "dU" is the normalized Haar measure on $R \otimes S \cong A \otimes B$.

A description of the calculation can be found in Appendix A. Let the eigenvalues of $(\mathcal{N} \otimes \bar{\mathcal{N}})(\Phi)$ be equal to $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{|A|^2}$. For a typical U, Lemmas III.3 and IV.1 together imply that

$$\sum_{j>1} \lambda_j^2 = \mathcal{O}\left(\frac{1}{|A|^2}\right). \tag{30}$$

Thus, aside from λ_1, the eigenvalues λ_j must be quite small. A typical eigenvalue distribution is plotted in Figure 1. If we define $\tilde{\lambda}_j = \lambda_j/(1-\lambda_1)$, then $\sum_{j>1} \tilde{\lambda}_j = 1$ and

$$H_1(\tilde{\lambda}) \geq H_2(\tilde{\lambda}) = -\ln \sum_{j>1} \tilde{\lambda}_j^2 = 2\ln|A| - \mathcal{O}(1). \tag{31}$$

An application of the grouping property then gives us a good lower bound on the von Neumann entropy:

$$H_1\big((\mathcal{N} \otimes \bar{\mathcal{N}})(\Phi)\big) = H_1(\lambda) = h(\lambda_1) + (1-\lambda_1)H_1(\tilde{\lambda}) = 2\ln|A| - \mathcal{O}(1), \tag{32}$$

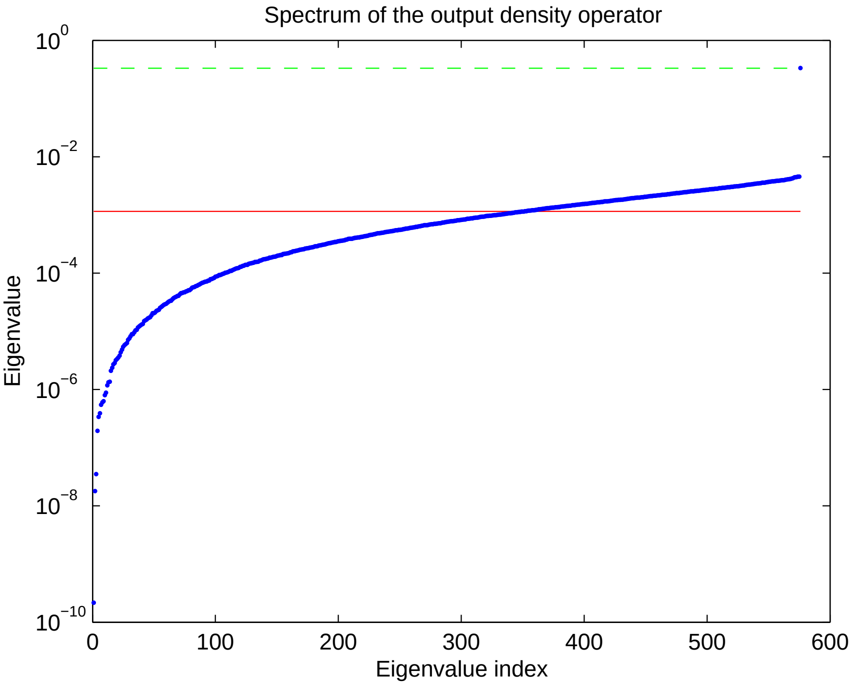


FIG. 1: Typical eigenvalue spectrum of $(\mathcal{N}\otimes\bar{\mathcal{N}})(\Phi)$ when $|R| = 3$ and $|A| = |B| = 24$. The eigenvalues are plotted in increasing order from left to right. The green dashed line corresponds to $|S|/(|A||B|) = 1/3$, which is essentially equal to the largest eigenvalue. The red solid line represents the value $(1-\frac{|S|}{|A||B|})/|A|^2 = 1/864$. If the density operator were maximally mixed aside from its largest eigenvalue, all but that one eigenvalue would fall on this line. While that is not the case here or in general, the remaining eigenvalues are nonetheless sufficiently small to ensure that the density operator has high von Neumann entropy.

where h is the binary entropy function. This entropy is nearly as large as it can be and, in particular, as large as $H^{\min}(\mathcal{N}) + H^{\min}(\bar{\mathcal{N}})$ according to Theorem IV.1 of [50], the von Neumann entropy version of Lemma III.2.

V. DISCUSSION

The counterexamples presented here demonstrate that the maximal p-norm multiplicativity conjecture and, equivalently, the minimum output p-Rényi entropy conjecture are false for all $1 < p \leq \infty$. The primary motivation for studying this conjecture was that it is a natural strengthening of the minimum (von Neumann) output entropy conjecture, which is of fundamental importance in quantum information theory. In particular, since the multiplicativity conjecture was formulated, most attempts to prove the minimum output entropy conjecture for special cases actually proved maximal p-norm multiplicativity and then took the limit as p decreases to 1. This strategy, we now know, cannot be used to prove the conjecture in general.

From that perspective, it would seem that the results in this paper cast doubt on the validity of the minimum output entropy conjecture itself. However, as we have shown, the examples explored here appear to be completely consistent with the conjecture, precisely because the von Neumann entropy is more difficult to perturb than the Rényi entropies of order $p > 1$. It is therefore still possible that the $p = 1$ conjecture could be demonstrated using subtle variants of p-norm multiplicativity such as exact or approximate multiplicativity in a channel-dependent interval $(1, 1 + \delta)$.

Another strategy that is still open would be to approach the von Neumann minimum output entropy via Rényi entropies for $p < 1$. It is possible that additivity holds there even as it fails for $p > 1$. That is not, unfortunately, a very well-informed speculation. With few exceptions [57], there has been very little research on the additivity question in the regime $p < 1$, even though many arguments can be easily adapted to this parameter region. (Eq. (13), for example, holds for all $0 < p$.) Unfortunately, since the time the examples presented here were first circulated, counterexamples for p close to 0 were also discovered [58], casting doubt on the conjecture for the whole set of Rényi entropies with $p < 1$. Indeed, as in the current paper, those examples are based on influencing a single eigenvalue of the output state of the tensor product channel; while here we increase the largest one, there the smallest is suppressed.

Thus, while it seems doubtful that the examples of channels presented here will have direct implications for the addivity of the minimum von Neumann entropy, we think that they are still very useful as a new class of test cases. Indeed, as we remarked earlier, our examples eliminate what had been the previously favoured route to the conjecture via the output p-norms.

As a final comment, while this paper has demonstrated that the maximal p-norm additivity conjecture fails for $p > 1$, all the counterexamples presented here have been nonconstructive. For the examples based on ϵ-randomizing maps, all the known explicit constructions (by Ambainis and Smith [59] or via iterated quantum expander maps [60, 61]) only give bounds in the 2-norm, which do imply bounds on the output p-norm but those are too weak to yield counterexamples to multiplicativity. Likewise, the counterexamples based on generic quantum channels rely on the existence of large subspaces containing only highly entangled states. Even when the entanglement is quantified using von Neumann entropy, in which case the existence of these subspaces was demonstrated in 2003 [50], not a single explicit construction is known. The culprit, as in many other related contexts [62], is our use of the probabilistic method. Since we don't have any *explicit* counterexamples, only a proof that counterexamples exist, it remains an open problem to "derandomize" our argument.

Acknowledgments

We would like to thank Frédéric Dupuis and Debbie Leung for an inspiring late-night conversation at the Perimeter Institute, Aram Harrow for several insightful suggestions, and Mary Beth Ruskai for discussions on the additivity conjecture. We also thank BIRS for their hospitality during the Operator Structures in Quantum Information workshop, which rekindled our interest in the additivity problem. PH was supported by the Canada Research Chairs program, a Sloan Research Fellowship, CIFAR, FQRNT, MITACS, NSERC and QuantumWorks. AW received support from the U.K. EPSRC, the Royal Society and the European Commission (project "QAP").

APPENDIX A: PROOF OF LEMMA IV.1

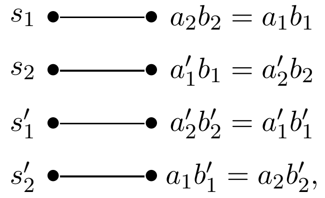
We will estimate the integral, in what is perhaps not the most illuminating way, by expressing it in terms of the matrix entries of U. Let $U_{s,ab} = {}^R\langle 0|^S\langle s|U|a\rangle^A|b\rangle^B$. Expanding gives

$$\int \mathrm{Tr}\left[\left((\mathcal{N}\otimes\bar{\mathcal{N}})(|\Phi\rangle\langle\Phi|)\right)^2\right] dU \tag{A1}$$
$$= \frac{1}{|S|^2}\sum_{\substack{a_1,a_2\\a_1',a_2'}}\sum_{\substack{b_1,b_2\\b_1',b_2'}}\sum_{\substack{s_1,s_2\\s_1',s_2'}}\int \bar{U}_{s_1,a_2b_2}\bar{U}_{s_2,a_1'b_1}\bar{U}_{s_1',a_2'b_2'}\bar{U}_{s_2',a_1b_1'}U_{s_1,a_1b_1}U_{s_2,a_2'b_2}U_{s_1',a_1'b_1'}U_{s_2',a_2b_2'}\,dU.$$

Following [63, 64], the non-zero terms in the sum can be represented using a simple graphical notation. Make two parallel columns of four dots, then label the left-hand dots by the indices (s_1, s_2, s_1', s_2') and the right-hand dots by the indices $\vec{v} = (a_2b_2, a_1'b_1, a_2'b_2', a_1b_1')$. Join dots with a solid line if the corresponding $\bar{U}$ matrix entry appears in Eq. (A1). Since terms integrate to a non-zero value only if the vector of U indices $\vec{w} = (a_1b_1, a_2'b_2, a_1'b_1', a_2b_2')$ is a permutation of the vector of $\bar{U}$ indices, a non-zero integral can be represented by using a dotted line to connect left-hand and right-hand dots whenever the corresponding U matrix entry appears in the integral.

Assuming for the moment that the vertex labels in the left column are all distinct and likewise for the right column, the integral evaluates to the Weingarten function $\mathrm{Wg}(\pi)$, where π is the permutation such that $w_i = v_{\pi(i)}$. For the rough estimate required here, it is sufficient to know that $\mathrm{Wg}(\pi) = \Theta\big((|A||B|)^{-4-|\pi|}\big)$, where $|\pi|$ is the minimal number of factors required to write π as a product of transpositions, and that $\mathrm{Wg}(e) = (|A||B|)^{-4}\big(1 + \mathcal{O}(|A|^{-2}|B|^{-2})\big)$ [65].

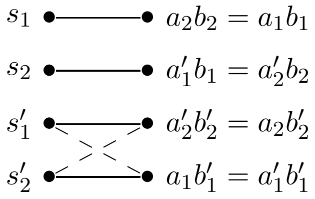
The dominant contribution to Eq. (A1) comes from the "stack" diagram



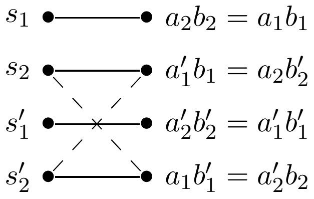
in which the solid and dashed lines are parallel and for which the contribution is positive and approximately equal to

$$\frac{1}{|S|^2}\sum_{\substack{a_1,a_2\\a_1',a_2'}}\sum_{\substack{b_1,b_2\\b_1',b_2'}}\sum_{\substack{s_1,s_2\\s_1',s_2'}}\delta_{a_1a_2}\delta_{b_1b_2}\delta_{a_1'a_2'}\delta_{b_1'b_2'}\,\mathrm{Wg(id)} = \frac{|S|^2}{|A|^2|B|^2}\left(1+\mathcal{O}(|A|^{-2}|B|^{-2})\right). \tag{A2}$$

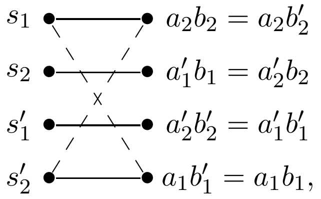
(The expression on the left-hand side would be exact but for the terms in which vertex labels are not distinct.) To obtain an estimate of Eq. (A1), it is then sufficient to examine the other terms and confirm that they are all of smaller asymptotic order than this. There are six diagrams representing transpositions, and their associated (negative) contributions are



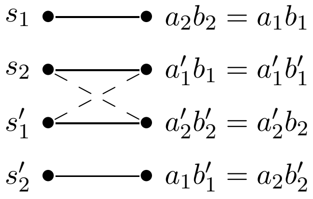
$\Theta(|S|^2|A|^{-4}|B|^{-2})$



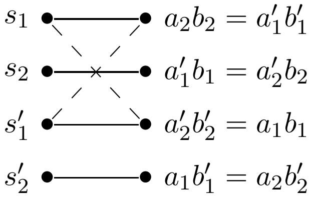
$\Theta(|S|^2|A|^{-4}|B|^{-4})$



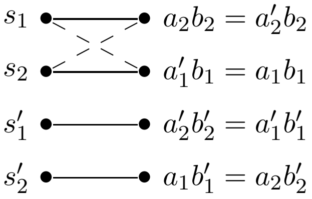
$\Theta(|S|^2|A|^{-2}|B|^{-4})$



$\Theta(|S|^2|A|^{-2}|B|^{-4})$



$\Theta(|S|^2|A|^{-4}|B|^{-4})$



$\Theta(|S|^2|A|^{-4}|B|^{-2})$.

For permutations π such that $|\pi| > 1$, the Weingarten function is significantly suppressed: $\mathrm{Wg}(\pi) = \mathcal{O}(|A|^{-6}|B|^{-6})$. Moreover, for a given diagram type, the requirement that $w_i = v_{\pi(i)}$ can only hold if at least two pairs of the indices $a_1, a_2, a_1', a_2', b_1, b_2, b_1', b_2'$ are identical. The contribution from such diagrams is therefore $\mathcal{O}(|S|^2|A|^{-4}|B|^{-2})$.

To finish the proof, it is necessary to consider integrals in which the vertex labels on the left- or the right-hand side of a diagram are not all distinct. In this more general case, choosing a set $\mathcal{C}$ of representatives for the conjugacy classes of the permutation group on four elements, the value of the integral can be written

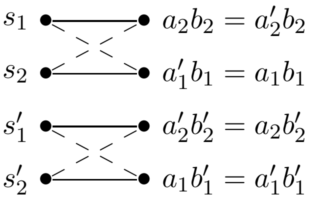
$$\sum_{c\in\mathcal{C}} N(c)\,\mathrm{Wg}(c), \tag{A3}$$

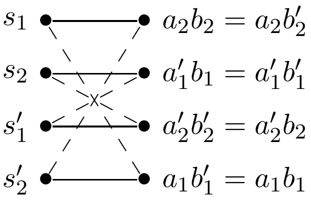
where

$$N(c) = \sum_{\substack{\sigma\in\mathcal{S}_4:\\ \vec{v}=\sigma(\vec{v})}} \sum_{\substack{\tau\in\mathcal{S}_4:\\ \vec{w}=\tau(\vec{w})}} \delta(\tau\pi\sigma \in c). \tag{A4}$$

These formulas have a simple interpretation. Symmetry in the vertex labels introduces ambiguities in the diagrammatic notation; the formula states that every one of the diagrams consistent with a given vertex label set must be counted, and with a defined dimension-independent multiplicity. Conveniently, our crude estimates have already done exactly that, ignoring the multiplicities. The only case for which we need to know the multiplicities, moreover, is for contributions to the dominant term, which we want to know exactly and not just up to a constant multiple.

We claim that in the sum (A1) there are at most $\mathcal{O}(|S|^4|A||B|^3)$ terms with vertex label symmetry. The total contribution for terms with vertex label symmetries τ and σ in which $|\tau\pi\sigma| \geq 1$ is therefore of size $\mathcal{O}(|S|^2|A|^{-4}|B|^{-2})$ and does not affect the dominant term. To see why the claim holds, fix a diagram type and recall that the requirement $w_i = v_{\pi(i)}$ for a permutation π can only hold if at least two pairs of the indices $a_1, a_2, a_1', a_2', b_1, b_2, b_1', b_2'$ are identical. Equality is achieved only when all the A indices or all the B indices are aligned, corresponding to the following two diagrams:





For the first diagram, using the fact that $|A| \leq |B| \leq |S|$, it is easy to check that imposing the extra constraint that either the top or bottom two S or AB vertex labels match singles at most $\mathcal{O}(|S|^4|A||B|^3)$ terms from Eq. (A1). Similar reasoning applies to the second diagram, but imposing the constraint instead on rows one and four, or two and three. For all other diagram types, at least four pairs of the indices $a_1, a_2, a_1', a_2', b_1, b_2, b_1', b_2'$ are identical. (The number of matching A

and B indices is necessarily even.) In a term for which the vertex labels are not all distinct, either a pair of S indices or a further pair of A or B indices must be identical. In the latter case, there must exist an identical A pair *and* an identical B pair among all the pairs. Again using $|A| \leq |B| \leq |S|$, there can be at most $\mathcal{O}(|S|^4|B|^3)$ such terms per diagram type, which demonstrates the claim.

We are thus left to consider integrals with vertex label symmetry and $N(e) \neq 0$ in Eq. (A3). If $N(e) = 1$, then our counting was correct and there is no problem. It is therefore sufficient to bound the number of integrals in which $N(e) > 1$. This can occur only in terms with at least 2 vertex label symmetries. Running the argument of the previous paragraph again, for the two diagrams with A or B indices all aligned, this occurs in at most $\mathcal{O}(|S|^4|B|^2)$ terms. For the rest of the cases, it is necessary to impose equality on yet another pair of indices, leading again to at most $\mathcal{O}(|S|^4|B|^2)$ terms. Since $\mathrm{Wg}(e) = \mathcal{O}(|A|^{-4}|B|^{-4})$, these contributions are collectively $\mathcal{O}(|S|^2|A|^{-4}|B|^{-2})$.

The bound on the error term in Eq. (29) arises by substituting the inequalities $|S| \leq |A||B|$ and $|A| \leq |B|$ into each of the estimates calculated above.

[1] B. Schumacher. Quantum coding. *Physical Review A*, 51:2738–2747, 1995.

[2] R. Jozsa and B. Schumacher. A new proof of the quantum noiseless coding theorem. *J. Mod. Opt.*, 41:2343–2349, 1994.

[3] J. Pierce. The early days of information theory. *IEEE Transactions on Information Theory*, 19(1):3–8, 1973.

[4] J. P. Gordon. Noise at optical frequencies; information theory. In P. A. Miles, editor, *Quantum electronics and coherent light; Proceedings of the international school of physics Enrico Fermi, Course XXXI*, pages 156–181, New York, 1964. Academic Press.

[5] A. S. Holevo. Information theoretical aspects of quantum measurements. *Probl. Info. Transm. (USSR)*, 9(2):31–42, 1973. Translation: Probl. Info. Transm. vol. 9, pp. 177-183, 1973.

[6] P. Hausladen, R. Jozsa, B. Schumacher, M. Westmoreland, and W. K. Wootters. Classical information capacity of a quantum channel. *Physical Review A*, 54:1869–1876, 1996.

[7] A. S. Holevo. The capacity of the quantum channel with general signal states. *IEEE Trans. Inf. Theory*, 44:269–273, 1998.

[8] B. Schumacher and M. D. Westmoreland. Sending classical information via noisy quantum channels. *Physical Review A*, 56:131–138, 1997.

[9] P. W. Shor. Equivalence of additivity questions in quantum information theory. *Communications in Mathematical Physics*, 246:453–472, 2004. arXiv:quant-ph/0305035.

[10] A. A. Pomeransky. Strong superadditivity of the entanglement of formation follows from its additivity. *Physical Review A*, 68(3):032317–+, September 2003. arXiv:quant-ph/0305056.

[11] K. M. R. Audenaert and S. L. Braunstein. On strong superadditivity of the entanglement of formation. *Communications in Mathematical Physics*, 246:443–452, 2004. arXiv:quant-ph/0303045.

[12] K. Matsumoto, T. Shimono, and A. Winter. Remarks on additivity of the Holevo channel capacity and of the entanglement of formation. *Communications in Mathematical Physics*, 246:427–442, 2004. arXiv:quant-ph/0206148.

[13] C. H. Bennett, D. P. DiVincenzo, J. A. Smolin, and W. K. Wootters. Mixed-state entanglement and quantum error correction. *Physical Review A*, 54:3824–3851, 1996. arXiv:quant-ph/9604024.

[14] P. M. Hayden, M. Horodecki, and B. M. Terhal. The asymptotic entanglement cost of preparing a quantum state. *Journal of Physics A: Mathematical and General*, 34:6891–6898, 2001. arXiv:quant-ph/0008134.

[15] G. Vidal, W. Dür, and J. I. Cirac. Entanglement cost of bipartite mixed states. *Physical Review Letters*, 89(2):027901–+, 2002. arXiv:quant-ph/0112131.

[16] K. Matsumoto and F. Yura. Entanglement cost of antisymmetric states and additivity of capacity of some quantum channels. *Journal of Physics A: Mathematical and General*, 37:L167–L171, 2004. arXiv:quant-ph/0306009.

[17] K. G. H. Vollbrecht and R. F. Werner. Entanglement measures under symmetry. *Physical Review A*, 64(6):062307–+, 2001. arXiv:quant-ph/0010095.

[18] C. King and M. B. Ruskai. Minimal entropy of states emerging from noisy quantum channels. *IEEE*

Transactions on Information Theory, 47(1):192–209, 2001. arXiv:quant-ph/9911079.

[19] S. Osawa and H. Nagaoka. Numerical experiments on the capacity of quantum channel with entangled input states. *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, E84-A(10):2583–2590, 2001. arXiv:quant-ph/0007115.

[20] G. G. Amosov, A. S. Holevo, and R. F. Werner. On some additivity problems of quantum information theory. *Probl. Inform. Transm.*, 36(4):25, 2000.

[21] G. G. Amosov and A. S. Holevo. On the multiplicativity conjecture for quantum channels. arXiv:math-ph/0103015, March 2001.

[22] C. King. Additivity for unital qubit channels. *Journal of Mathematical Physics*, 43(10):4641–4643, 2002. arXiv:quant-ph/0103156v1.

[23] A. Fujiwara and T. Hashizumé. Additivity of the capacity of depolarizing channels. *Physics Letters A*, 299:469–475, July 2002.

[24] C. King. The capacity of the quantum depolarizing channel. *IEEE Transactions on Information Theory*, 49(1):221–229, 2003. arXiv:quant-ph/0204172.

[25] A. S. Holevo. Quantum coding theorems. *Russ. Math. Surv.*, 53:1295–1331, 1998.

[26] C. King. Maximization of capacity and p-norms for some product channels. arXiv:quant-ph/0103086, 2001.

[27] P. W. Shor. Additivity of the classical capacity of entanglement-breaking quantum channels. *Journal of Mathematical Physics*, 43:4334–4340, 2002. arXiv:quant-ph/0201149.

[28] I. Devetak and P. W. Shor. The capacity of a quantum channel for simultaneous transmission of classical and quantum information. *Communications in Mathematical Physics*, 256:287–303, June 2005.

[29] C. King, K. Matsumoto, M. Nathanson, and M. B. Ruskai. Properties of conjugate channels with applications to additivity and multiplicativity. arXiv:quant-ph/0509126, 2005.

[30] J. Cortese. Holevo-Schumacher-Westmoreland channel capacity for a class of qudit unital channels. *Physical Review A*, 69(2):022302–+, 2004.

[31] N. Datta, A. S. Holevo, and Y. M. Suhov. A quantum channel with additive minimum output entropy. arXiv:quant-ph/0403072, 2004.

[32] M. Fukuda. Extending additivity from symmetric to asymmetric channels. *Journal of Physics A: Mathematical and General*, 38:L753–L758, November 2005. arXiv:quant-ph/0505022.

[33] A. S. Holevo. Additivity of classical capacity and related problems. Available online at: http://www.imaph.tu-bs.de/qi/problems/10.pdf, 2004.

[34] A. S. Holevo. The additivity problem in quantum information theory. In *Proceedings of the International Congress of Mathematicians, Madrid, Spain, 2006*, Publ. EMS, pages 999–1018, Zurich, 2007.

[35] C. King and M. B. Ruskai. Comments on multiplicativity of maximal p-norms when p= 2. *Quantum Information and Computation*, 4:500–512, 2004. arXiv:quant-ph/0401026.

[36] C. King, M. Nathanson, and M. B. Ruskai. Multiplicativity properties of entrywise positive maps. *Linear algebra and its applications*, 404:367–379, 2005. arXiv:quant-ph/0409181.

[37] A. Serafini, J. Eisert, and M. M. Wolf. Multiplicativity of maximal output purities of Gaussian channels under Gaussian inputs. *Phys. Rev. A* , 71(1):012320–+, January 2005.

[38] V. Giovannetti and S. Lloyd. Additivity properties of a Gaussian channel. *Physical Review A*, 69:062307, 2004. arXiv:quant-ph/0403075.

[39] I. Devetak, M. Junge, C. King, and M. B. Ruskai. Multiplicativity of completely bounded p-norms implies a new additivity result. *Communications in Mathematical Physics*, 266:37–63, August 2006.

[40] S. Michalakis. Multiplicativity of the maximal output 2-norm for depolarized Werner-Holevo channels. arXiv:0707.1722, 2007.

[41] R. F. Werner and A. S. Holevo. Counterexample to an additivity conjecture for output purity of quantum channels. *Journal of Mathematical Physics*, 43:4353–4357, 2002. arXiv:quant-ph/0203003.

[42] R. Alicki and M. Fannes. Note on multiple additivity of minimal Renyi entropy output of the Werner-Holevo channels. *Open Systems and Information Dynamics*, 11(4):339–342, 2005. arXiv:quant-ph/0407033.

[43] N. Datta. Multiplicativity of maximal p-norms in Werner-Holevo channels for $1 < p < 2$. arXiv:quant-ph/0410063, 2004.

[44] V. Giovannetti, S. Lloyd, and M. B. Ruskai. Conditions for multiplicativity of maximal p -norms of channels for fixed integer p. *Journal of Mathematical Physics*, 46:042105, 2005. arXiv:quant-ph/0408103.

[45] A. Winter. The maximum output p-norm of quantum channels is not multiplicative for any $p>2$.

arXiv:0707.0402, 2007.

[46] P. Hayden. The maximal p-norm multiplicativity conjecture is false. arXiv.org:0707.3291, 2007.

[47] P. Hayden, D. Leung, P. W. Shor, and A. Winter. Randomizing Quantum States: Constructions and Applications. *Communications in Mathematical Physics*, 250:371–391, 2004.

[48] G. Aubrun. On almost randomizing channels with a short Kraus decomposition. arXiv.org:0805.2900v2, 2008.

[49] V. I. Paulsen. *Completely bounded maps and dilations*. Longman Scientific and Technical, New York, 1986.

[50] P. Hayden, D. W. Leung, and A. Winter. Aspects of generic entanglement. *Communications in Mathematical Physics*, 265:95–117, 2006. arXiv:quant-ph/0407049.

[51] C.H. Bennett, P. Hayden, D.W. Leung, P.W. Shor, and A. Winter. Remote preparation of quantum states. *IEEE Transactions on Information Theory*, 51(1):56–74, Jan. 2005. arXiv:quant-ph/0307100.

[52] S. Geman. A Limit Theorem for the Norm of Random Matrices. *Annals of Probability*, 8(2):252–261, 1980.

[53] I. M. Johnstone. On the distribution of the largest eigenvalue in principal components analysis. *Annals of Statistics*, 29(2):295–327, 2001.

[54] K. R. Davidson and S. J. Szarek. Local Operator Theory, Random Matrices and Banach Spaces. In W. B. Johnson and J. Lindenstrauss, editors, *Handbook of the Geometry of Banach Spaces, Vol. I*, chapter 8, pages 317–366. Elsevier, 2001.

[55] M. Ledoux. *The concentration of measure phenomenon*, volume 89 of *Mathematical Surveys and Monographs*. American Mathematical Society, 2001.

[56] C. King. Maximal p-norms of entanglement breaking channels. *Quantum Information and Computation*, 3(2):186–190, 2003. arXiv:quant-ph/0212057.

[57] M. M. Wolf and J. Eisert. Classical information capacity of a class of quantum channels. *New Journal of Physics*, 7:93–+, 2005. arXiv:quant-ph/0412133.

[58] T. Cubitt, A. W. Harrow, D. Leung, A. Montanaro, and A. Winter. Counterexamples to additivity of minimum output p-Rényi entropy for p close to 0. arXiv.org:0712.3628v2, 2007.

[59] A. Ambainis and A. Smith. Small pseudo-random families of matrices: Derandomizing approximate quantum encryption. In *Proc. RANDOM*, LNCS 3122, pages 249–260. Springer, 2004. arXiv.org:quant-ph/0404075.

[60] A. Ben-Aroya and A. Ta-Shma. Quantum expanders and the quantum entropy difference problem. arXiv.org:quant-ph/0702129, 2007.

[61] M. B. Hastings. Random unitaries give quantum expanders. *Physical Review A*, 76(3):032315–+, 2007. arXiv:0706.0556.

[62] D. Pérez-García, M. M. Wolf, C. Palazuelos, I. Villanueva, and M. Junge. Unbounded Violation of Tripartite Bell Inequalities. *Communications in Mathematical Physics*, 279(2):455–486, 2008. arXiv:quant-ph/0702189.

[63] S. Aubert and C. S. Lam. Invariant integration over the unitary group. *Journal of Mathematical Physics*, 44:6112–6131, 2003. arXiv:math-ph/0307012.

[64] S. Aubert and C. S. Lam. Invariant and group theoretical integrations over the U(n) group. *Journal of Mathematical Physics*, 45:3019–3039, 2004. arXiv:math-ph/0405036.

[65] B. Collins and P. Śniady. Integration with respect to the Haar measure on unitary, orthogonal and symplectic group. *Communications in Mathematical Physics*, 264:773–795, 2006. arXiv:math-ph/0402073.